Biochimica et Biophysica Acta 1809 (2011) 700–707

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

# Drosha processing controls the specificity and efficiency of global microRNA expression☆

Yong Feng [a], Xiaoxiao Zhang [a], Qingfeng Song [b], Tongbin Li [b], Yan Zeng [a,c,*]

[a] Department of Pharmacology, University of Minnesota, Minneapolis, MN 55455, USA
[b] Biolead.org Research Group, LC Sciences, Houston, TX 77054, USA
[c] Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

*Article history:*
Received 13 April 2011
Received in revised form 19 May 2011
Accepted 20 May 2011
Available online 13 June 2011

*Keywords:*
miRNA
Pri-miRNA processing
Drosha
miRNA expression
RNA secondary structure

ABSTRACT

microRNAs (miRNAs) are a large family of approximately 22-nucleotide-long RNAs that regulate gene expression. They are first transcribed as long, primary transcripts, which then undergo a series of processing steps to generate the single-stranded, mature miRNAs. Here, we showed that Drosha cleaved hundreds of human primary miRNA transcript substrates with different efficiencies in vitro. The differential Drosha susceptibility of the primary miRNA transcripts significantly correlated with the expression of the corresponding, mature miRNAs in vivo. Conserved miRNAs were more efficiently expressed in vivo, and their primary transcripts were also better Drosha substrates in vitro. Combining secondary structure prediction and statistical analyses, we identified features in human primary miRNA transcripts that predisposed miRNAs to efficient Drosha processing in vitro as well as to better expression in vivo. We propose that the selectivity of Drosha action contributes greatly to the specificity and efficiency of miRNA biogenesis. Moreover, this study serves as an example of substrate specificity of a biochemical reaction regulating gene expression at a global scale in vivo. This article is part of a Special Issue entitled: MicroRNA's in viral gene regulation.



## 1. Introduction

miRNAs are a class of non-coding RNAs that regulate target gene expression and control a wide range of biological processes [1]. They are ~22 nucleotide (nt) long and embedded within one arm of a hairpin structured transcript, with rare exceptions [2,3]. This feature enables the prediction and classification of miRNAs [4]. Significant gaps, however, remain in our understanding of miRNA biogenesis, and our ability to correctly predict or identify miRNAs is limited. Many miRNA mining algorithms predict miRNA genes to number up to hundreds of thousands in complex genomes (e.g., [5–7]), notably more than the miRNAs currently deposited in the miRBase [8]. This is not surprising, because relatively little is known about the determinants of genuine miRNA transcripts, while the genomes conceivably encode tens of millions of hairpin RNA elements. Indeed, more and more candidate miRNAs are being identified by deep sequencing. Nonetheless, it has also been reported that merely several hundred miRNA genes are expressed in mammals, many of them only rarely [9,10].

A miRNA is first transcribed as part of a long primary transcript, or pri-miRNA [11]. The pri-miRNA is cleaved by an RNase called Drosha, complexed with its regulatory subunit DGCR8 in mammals, to liberate a hairpin precursor, or pre-miRNA, of ~60–70 nt [12–16]. The pre-miRNA is then exported to the cytoplasm by Exportin5 [17–19] and digested by the RNase Dicer to produce an ~22 basepair miRNA duplex intermediate [20–23]. An Argonaute protein binds to the duplex and finally selects the mature, single-stranded miRNA. A small minority of miRNAs can be generated independent of Drosha or Dicer [24–28]. Given such a huge pool of potential pri-miRNAs but apparently only a tiny fraction producing detectable, mature miRNAs, we hypothesized that, in addition to the obvious regulation at the transcriptional level, the miRNA processing machinery can also distinguish whether a potential RNA is a good substrate or a poor one; i.e., miRNA processing intrinsically serves a checkpoint or regulatory function. Indeed, a recent paper suggested that processing can severely limit RNA expression from a library of 20,000 short hairpin RNAs [29]. Our current work investigated the function of the Drosha/DGCR8 holoenzyme, or Drosha in short hereafter, because it initiates the irreversible miRNA processing and conceivably determines both how fast a pre-miRNA is produced and the fate of its substrates, as Drosha cleavage products ultimately undergo additional processing or degradation. We asked these questions: at the global scale, how is miRNA expression regulated? If mammals really encode hundred of thousands of miRNA genes, why aren't many produced at a significant level? And lastly, can we identify the structural

*Abbreviations:* miRNA, microRNA; pri-miRNA, primary microRNA transcript

☆ This article is part of a Special Issue entitled: MicroRNA's in viral gene regulation.

* Corresponding author at: Department of Pharmacology, University of Minnesota, 6-120 Jackson Hall, 321 Church St SE, Minneapolis, MN 55455, USA. Tel.: +1 612 625 4479; fax: +1 612 625 8408.
*E-mail address:* zengx033@umn.edu (Y. Zeng).

1874-9399/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagrm.2011.05.015

features that distinguish miRNA transcripts from other RNAs and that determine the efficiency of miRNA processing and expression in vivo?

## 2. Materials and methods

### 2.1. Drosha cleavage assay

DNA templates for RNA synthesis were generated by PCR with a primer containing the T7 promoter sequence, using human genomic DNA (Clontech) as template. The pri-miRNA substrates were prepared by in vitro transcription (Promega) in the presence of [α-$^{32}$P] CTP. The RNAs contained the pre-miRNA moiety flanked by ~25 nt extra sequences at both sides, for we and others had demonstrated that such pri-miRNAs contained all the essential elements for Drosha cleavage in vitro [30,31]. We also tested a number of pri-miRNAs with 50–200 nt flanking regions and got similar results (data not shown). Size markers ϕX174 DNA/HinfI (Promega) were labeled at their 5′-ends with [γ-$^{32}$P] ATP by T4 polynucleotide kinase (New England BioLabs). Drosha holoenzyme was prepared, used at ~1 ng/μl, mixed with approximately equal amount of radioactive RNA substrates, and incubated at 37 °C for 45–60 min [32]. End-point analysis was performed not only because of the sheer number of the pri-miRNA substrates but also because we had found that both time-course and end-point experiments yielded the same conclusions ([32]; data not shown). Because we used Spearman rank correlation analysis (see Section 2.5 below), the rank of the cleavage efficiencies is more important than the actual values, which also was easier to satisfy by end-point studies. After gel electrophoresis, data were analyzed using a PhosphorImager. Cleavage efficiency was calculated as the intensities of predicted products (the ~60–70 nt pre-miRNA and the ~25 nt flanking RNA) divided by the intensities of the products and the remaining, full-length pri-miRNA. The pri-let-7a substrate was included in every experiment so that processing efficiencies of all the other pri-miRNAs could be compared to that of pri-let-7a, which was set as 100. Pri-let-7a was not necessarily cleaved at the same rate in every assay due to experimental variations, e.g., probes were prepared and used on different days, reaction volume might not be exactly the same each time, and different activities of Drosha might be added due to pipetting or different batches of enzyme preparations. Nevertheless, pri-let-7a was chosen as the control because it was one of the most efficiently cut RNA in every experiment (so that one could make sure that the assay worked), and other pri-miRNAs, tested on different days, gave similar, relative efficiencies when normalized to pri-let-7a.

### 2.2. Cell culture and transfection

293T cells were maintained at 37 °C and 5% $CO_2$ in Dulbecco's modified Eagle's medium with 10% fetal bovine serum. Plasmids that overexpress human miRNAs were constructed by amplifying typically 500 bp DNA from human genomic DNA and inserting it after a cytomegalovirus promoter [33]. Equal amounts of different plasmids were transfected individually into 293T cells using Lipofectamine 2000 (Invitrogen). Total RNA was isolated using Trizol reagent (Invitrogen).

### 2.3. Northern blotting

Northern blot analysis was performed as described [33]. Sequences of the oligonucleotide probes are: for miR-7 detection, 5′-ACAACAAAATCACTAGTCTTCCA-3′; miR-105, 5′-ACCACAGGAGTCTGAGCATTTGA-3′; miR-125b, 5′-TCACAAGTTAGGGTCTCAGGGA-3′; miR-504, 5′-GATAGAGTGCAGACCAGGGTCT-3′; miR-634, 5′-GTCCAAAGTTGGGGTGCTGGTT-3′; miR-765, 5′-GCATCACCTTCCTTCTCCTCCA-3′; and U6 snRNA control: 5′-ACGAATTTGCGTGTCATCCTTGCG-3′. Results were analyzed by autoradiography or using a Storm 840 PhosphorImager (GE Healthcare).

### 2.4. Real-time PCR

Total RNAs were first polyadenylated and then reverse transcribed with an oligo(dT)-linker primer using the miScript system (Qiagen), and cDNA amplified by PCR in 40 cycles of 95 °C 10 s, 60 °C 45 s, on a Stratagene Mx3005P machine. U6 served as the internal control.

### 2.5. Statistics and secondary structure prediction

GraphPad Prism 5.0 (GraphPad Software) and SPSS 13.0 (IBM) were used for Spearman rank correlation and Mann–Whitney U tests (two-tailed). The p-value of Spearman rank correlation was calculated with a permutation test. miRNA genomic sequence, miRNA family, species conservation, and cluster information was retrieved from miRBase. The expression data of a miRNA were calculated by adding the sequence reads for the miRNA and its miRNA* from all the tissues and cell lines [9,10]. The combination of a wide variety of tissues and cell lines conceivably, partially reduced cell-specific effects due to, e.g., differential gene transcription and the expression of proteins that regulate the processing of specific miRNAs [11,34]. For secondary structure prediction, the actual pri-miRNA substrates (listed in Supplementary Table S2) were folded using Mfold version 3.2, under the default condition of 37 °C and 1 M NaCl [35,36]. For all the other pri-miRNAs, their folding used 5′ and 3′, 25-nt-long extensions beyond the pre-miRNA moiety. For the multiple genes that produce the same miRNA, e.g., let-7a-1, let-7a-2, and let-7a-3, only one (let-7a-1) was used for prediction and correlation. ΔG of the terminal loop region was calculated from the pri-miRNA structural prediction. The terminal loop region starts from the first nucleotide after the 3′ end of the actual or predicted miRNA or miRNA* at the 5′ arm of the hairpin and ends at its corresponding nucleotide at the 3′ arm. For ΔG of the miRNA duplex region, a pre-miRNA was folded using Mfold, and its ΔG subtracted by that of the terminal loop region. ΔG for the proximal and distal domains of the pre-miRNA-flanking region was analogously computed. The proximal domain contains 12 nt extensions beyond the 5′ and 3′ ends of a pre-miRNA. Typically the most stable conformations were analyzed. Exceptions were allowed to ensure that the predicted miRNA duplex moiety formed usually 2 nt, but at least 1 nt 3′ overhangs within a pri-miRNA. There are inherent heterogeneities in the 5′ and 3′ ends of mature miRNAs, and the secondary structure predictions can only be estimates. For secondary structure prediction shown in Fig. 5B, pri-miRNAs (listed in Supplementary Table S2) containing a larger distal domain were folded using Mfold, and the numbers of base pairs within the 12-nt-long, D1, F1, F2, F3, and F4 segments were divided by 12 to obtain the basepairing ratios.

## 3. Results

### 3.1. Human pri-miRNAs are processed with different efficiencies by Drosha in vitro

The processing of only a handful of pri-miRNAs by Drosha had been dissected (e.g., [12,15,30,31,33]). Here, we decided to take an unbiased, global approach. There were 740 human miRNA genes in the miRBase Release 14 [8]. For the first 550 miRNA genes we chose a representative from every miRNA family, e.g., let-7a-1 (let-7a in short) from the let-7 family (Supplementary Table S1). For the remaining miRNAs we randomly selected only 12 for analysis because then-finished Drosha assays suggested that they might not be cleaved efficiently (Supplementary Table S1; see below). We also excluded well-known Drosha-independent miRNAs [24–26]. In vitro Drosha processing assays were performed on the chosen pri-miRNAs (see Section 2.1 above). Excluding subjects that could not be amplified or produced RNA not suitable for our assays, we obtained reliable processing data for 247 pri-miRNAs (Supplementary Table S1). Results of three representative experiments performed on separate days are shown in Fig. 1. We found